Expression Robust 3D Face Recognition by Matching Multi-component Local Shape Descriptors on the Nasal and Adjoining Cheek Regions

Jiangning Gao, Adrian N Evans
Department of Electronic and Electrical Engineering, University of Bath, Bath, UK

Abstract— **This paper proposes a novel local depth and surface normals descriptor to explore the discriminative features on the nasal surface and the adjoining cheek regions for expression robust 3D face recognition. After preprocessing the 3D face data, landmarks located on the perimeter of a triangular region covering the nose and adjoining parts of the cheeks are accurately detected. Inspired by Local Binary Patterns, local shape differences for 3D points on a set of horizontal curves joining selected landmarks provide a novel representation of the local shape information. A further analysis of the discriminatory power of each patch shows that the adjoining regions have the potential to produce good recognition performance. Using the FRGC and Bosphorus databases, the performance of the proposed descriptor is evaluated on diverse patches, scales and for four components, one from the depth and three from the surface normals. Results show that the new local shape descriptor performs well at representing the shape information on a relatively large scale. On the basis of this descriptor, a relatively small set of features extracted from the nasal and adjoining cheek regions produce a R_1RR of 97.76% and an EER of 1.32%. The adjoining cheek regions demonstrate a high discriminatory power and provide a useful new addition to 3D face biometrics.**

I. INTRODUCTION

Overcoming variations in human expressions, poses and occlusions are some of the most challenging topics in face recognition. If only 2D captures are available it is difficult to accommodate these variations or to extract stable features to represent the human face. In the last decade, 3D face recognition has attracted much interest due to the increasing feasibility of 3D image acquisition and newly developed recognition algorithms. Using 3D data facilitates the correction of pose and can help address the problems caused by occlusions. However, variations in expression which lead to muscle movements and deform the face surface still present challenges to recognition systems using 3D data.

One way to address these problems is to select relatively stable structures and patches on the facial surface and use these to explore its discriminative features to design expression invariant face recognition algorithms. For example, the nasal region, which is more consistent over universal expressions and also invariant to the majority of occlusions caused by hair, hands and scarves [1-5], is widely used in many face recognition algorithms. Mian *et al*. [3] found the forehead is also relatively invariant under basic expressions. However, the forehead is sensitive to occlusion by human hair and it can be hard to extract effective features for recognition from it.

Although the experiments in [3] demonstrated that the cheek region is more susceptible to expression variations than the eyes, forehead and the nose, those parts of cheeks adjoining the nose, between the cheek bones and nasal bridge, potentially contain rich discriminative features which can be used as part of an expression robust recognition scheme. Chang *et al*. produced a good recognition performance by matching multiple overlapping 3D region of the nose and its surroundings [1], demonstrating that such regions have much potential for improving expression robustness. Wang *et al*. also explored the nose and its surrounding region to build a more efficient classifier [4] and Ballihi *et al*. found that the most relevant circular curves are located around nasal region [6]. These results demonstrate that the nasal shape makes a significant contribution to face recognition, as it is stable to facial expression. Recently, Gao *et al*. explored the use of the "cheek/nose" region, defined by a triangular structure containing the nose and adjoining parts of the cheek, to build an effective pattern rejector [7]. This work showed the potential of extending the nasal region to include adjoining parts of the cheeks for expression robust 3D face recognition.

Although many methods can be used for extracting features from the 3D face, depth information is commonly used because of its efficiency. One of the most popular approaches is to draw curves between extracted landmarks. Drira *et al*. first manually located the nose tip and then used a set radial curves emanating from the nose tip to develop a Riemannian framework for analyzing shapes of full facial surfaces [8]. Although this approach overcomes some of the significant problems in face recognition and produces a high recognition performance, utilizing information from the whole face region is challenging and also requires a non-trivial preprocessing step to normalize the whole face and correct for occlusions. Ballihi *et al*. extracted a set of geometric features including circular and radial curves [6]. However, the location of the curves covered some expression sensitive regions, presenting problems for expression robust recognition and, the nose tip also had to be detected manually. Emambakhsh *et al*. automatically detected four significant landmarks, the nose tip, the saddle and the two alar, and used these to define a set of 28 nasal curves that were used for expression robust recognition [2]. In [7], Gao *et al*. only used four curves, one curve from the nose tip to the saddle and three horizontal curves across the cheek/nose region, to form a low dimensionality pattern rejector that is robust to expression variations. Continued use of the nose region supports the hypothesis that it is relatively stable under expressions and less affected by occlusions.

To explore the recognition potential of local regions many works employ features such as Local Binary Patterns [9]. Li *et al*. successfully used multi-scale and multi-component local

normal patterns to find the expression robust patches on the whole 3D face, showing the discriminatory power of both Local Binary Patterns and surface normals [10]. However, their use of three scales and three components on the whole face results in high dimensional feature set for matching and can also suffer more from occlusions and expression variations. Therefore, exploring the local normal features on relatively rigid regions of the face is an attractive area of investigation.

Most 3D face recognition algorithms are focused on using depth information to select features and produce good recognition performance. In addition to depth, the surface normal of each point determines the orientation of the surface and contains information on local shape variations. For example, Zafeiriou *et al*. obtain the surface normal information from the photometric stereo and use it as a normal face for face recognition [11]. In comparison, Li *et al*. use the surface normal calculated through 3D point cloud data which is captured from laser scanner and demonstrate a good recognition performance [10]. These works provide motivation for the fusion of depth and surface normal to exploit more discriminative features for 3D face recognition.

This paper further investigates the use of the nasal and immediately adjoining cheek regions as a biometric. It proposes a novel local shape descriptor to find expression robust, discriminative features on both depth and normal maps. This descriptor is calculated between points from a set of three horizontal lines and their eight neighbors at a fixed separation. Evaluating the features over various scales demonstrates that this local shape descriptor provides a novel way to represent the shape changes on the face surface and produce a better recognition performance than using depth or normal values independently. In addition, a further analysis of the discriminatory power using 24 patches on the nose and adjoining cheek regions is also evaluated, see section IV. A discriminatory power map is produced, showing the potential of the region immediately adjoining the nose as a biometric. On the basis of the resulting discriminative map of the depth and three normal components, an extension of this new local shape descriptor is applied to the nine patches on the selected region. The recognition performance demonstrates that features extracted on this combined region by the proposed local shape descriptor can be developed as an expression robust 3D face recognition algorithm.

II. 3D Face Databases, Preprocessing and Landmarks Localization

A. 3D Face Database

For the preprocessing, feature extraction and performance evaluation the FRGC [12] and Bosphorus [13] databases are employed. The FRGC database contains a large number of captures with many variations in expression, pose and some basic occlusions. It is considered by some researchers to be the most challenging database for face recognition [14]. For Experiment 3 in [12], the 3D captures in the Spring2003 folder, consisting of 943 captures from 275 subjects, are used for training and the other two folders with more variations, Fall2003 and Spring2004, consisting of 4007 captures from 466 subjects, named FRGC v.2, are used for validation. To extract more effective discriminative features for expression robust face recognition, those expressions that only cause minor changes to the face surface were considered as neutral. The resulting manual division produced neutral and non-neutral sets of size 3396 and 1554, respectively. Although nearly all the captures in Spring2003 are neutral, the training set is still defined as "All". The testing set is either considered as "All" or divided into "Neutral" and "Non-neutral" and these three scenarios, "All vs All", "All vs Neutral" and "All vs Non-neutral", are used in the following experiments.

The Bosphorus database possesses many kinds of action units from the Facial Action Coding System, which can represent different changes on the face surface caused by human muscle movement, and also includes the basic universal expressions. Nearly all the captures contain different degrees or types of expressions and thus evenly dividing them into the training and testing sets guarantees that all kinds of expressions are evenly distributed in both subsets. This scenario is termed "All vs All" and has 1346 training and 1300 test captures, respectively. As this work aims to find the most discriminative and expression robust features, only those captures without large pose variations or occlusions are considered.

B. Preprocessing and Landmarks Localization

This work employs a four-step preprocessing scheme. First, the nose tip is detected by finding the largest convex region on the face and then the main face region is found by intersecting its surface with a sphere centered on the nose tip. Following Mian *et al*. [3], the radius of the sphere was set to 80mm although this is not critical to the overall performance as subsequent features are only extracted from the nasal region. Spike removal [3], hole filling and smoothing filters are sequentially applied to the cropped face to reduce the effects of these three artefacts. Finally, an improved Principle Component Analysis based pose correction algorithm is used to address the pose variations [2, 3].

After the preprocessing and normalization of the 3D captures, 16 landmarks are detected on the perimeter of a triangular area containing the nasal region, as shown in Fig. 1. The nose tip (9) and the saddle (1) are detected using a curvature-based approach [2] then, following [7], the extremities of the region containing the nose and the adjoining parts of the cheek (landmarks 5 and 13) are defined by extending a line horizontally from the nose tip by a fixed pixel distance. This was found to be more effective than using a proportional distance [7]. The remaining landmarks are found by dividing the lines (1 to 5, 5 to 9, 9 to 13 and 13 to 1) into four equal segments and the positions of the resulting landmarks are shown in Fig. 1.

After the landmarks are localized, surface features are extracted to first defining planes connecting any two landmarks, which are perpendicular to the x-y plane. This results in a series of curves on the face surface [2, 7]. Feature sets can then be found by normalizing the curves and then resampling them to a fixed number of points using bicubic interpolation. Gao *et al*. found that using 15 points from horizontal lines was an effective approach to capturing the discriminative features [7].

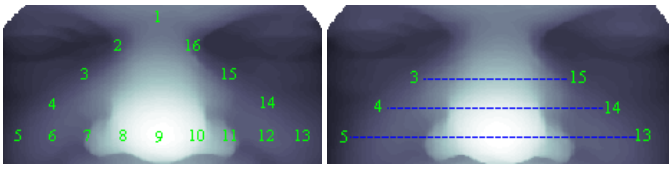


Fig. 1. Landmarks and horizontal curves extracted on the cheek/nose region.

III. An Expression Robust Multi-component Local Shape Descriptor

A. Building Local Shape Descriptors by Horizontal Curves

As an alternative to using the direct depth values of points from curves on and around the nasal region, this paper presents a novel representation for describing the shape information changes by computing depth differences, in a similar manner to Local Binary Patterns [9]. From the points extracted on the three horizontal curves shown in Fig. 1, a series of windows (the white quadrilaterals in Fig. 2) are used to compute the depth differences between the central points (red points) on the middle curve and the surrounding points (white points) on the top, middle and bottom curves. Specifically, all the even points on the middle curve are considered as the central points of 3×3 windows and the depth differences with their eight neighbours are stored as D_w= [dd_1, dd_2, dd_3, dd_4, dd_5, dd_6, dd_7, dd_8]. Therefore, the feature vector for each capture can be defined as (D_{w2}, D_{w4}, D_{w6}, … , $D_{w(i-1)}$), in which i is the number of points on each curve. In the experiments, i is set to 15 so there are 15 points on each curve and a total of 7 windows for calculation, giving a feature set size of 8×7 = 56. For comparison, a feature set of size 15×3 = 45, consisting of the depth values at the 15 points on each curve is used.

The recognition performance is measured by the equal error rate (EER) of the receiving operating characteristic (ROC) curve for verification and the rank one recognition rate (R_1RR) of the cumulative match characteristic (CMC) curve for identification. Kernel Fisher's Analysis (KFA) [15] with the polynomial kernel is used to calculate the similarity between the gallery and probe captures and the cosine distance is used in the feature matching. A comparison of recognition results between this new local shape descriptor and direct depth features is shown in TABLE I. The results show that using the depth differences performs better for all scenarios, in particular, produce a R_1RR that is about 9% higher than the linear depth features for the "All vs Non-neutral" scenario. This demonstrates that this local shape descriptor has the potential to produce a better recognition performance than using the depth values directly, especially when the captures contain different kinds of expressions.

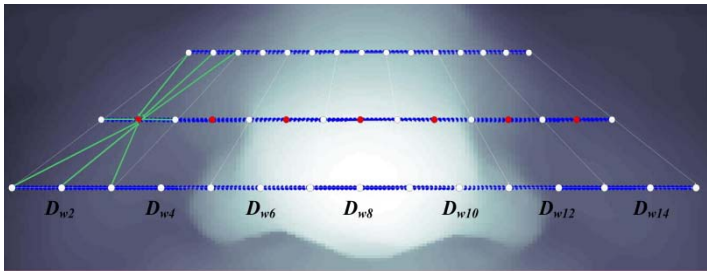


Fig. 2. Depth difference calculation using three horizontal curves.

TABLE I. R_1RR and EER performance for depth difference and direct depth features extracted from the horizontal curves in Fig. 2

	Scenarios	Depth Difference		Direct Depth	
		R_1RR	**EER**	**R_1RR**	**EER**
Bosphorus[13]	All vs All	88.98%	4.93%	85.82%	5.10%
Experiment 3 in FRGC [12]	All vs All	83.53%	5.34%	79.32%	5.32%
	All vs Neutral	90.50%	2.88%	89.02%	3.42%
	All vs Non-neutral	70.91%	7.02%	61.85%	8.56%
Features size of each capture		8×7 = 56		15×3 = 45	

B. Local Depth and Normal Features Extraction on the Nose and Adjoining Cheek Regions

The results in the previous section support the conclusions of previous researchers, that the nose and adjoining regions of the cheek provide a good source of discriminative features [1, 4, 6, 7]. In addition, in comparison to using the depth directly, using the proposed local shape descriptor to build the feature vector by calculating the depth differences is an effective and efficient method to describe the shape of the nasal region. In the previous section, the depth differences were calculated over a relatively large scale and in this section the discriminative features for expression invariant 3D face recognition are further explored by computing the depth difference more locally on the nose and adjoining parts of the cheek.

With reference to Fig. 3, the local depth descriptor is found using the points on the three horizontal curves shown in Fig. 2. Again, the curves are resampled to 15 points using bicubic interpolation and all the even points on the curves (marked as red points in Fig. 3) are taken as the central points of the 3×3 windows for which the depth differences are found. For each central point the eight neighboring points are given by the left and right points on the same curve and the three points directly above and below from curves at a vertical separation D, shown in blue in Fig. 3. D is measured in pixels at the resolution of the original 2D maps produced from the 3D point cloud. To allow the scale for different regions to be explored independently, the vertical distances D_1, D_2 and D_3 are used for the top, middle and bottom curves, respectively. For the seven central points on each of the three curves the resulting depth features produce a feature vector of size 8×7×3 = 168.

Fig. 4 plots the R_1RR achieved by the depth descriptor for the top, middle and bottom curves using different values of distance D for both the FRGC and Bosphorus databases. In general, the depth descriptors extracted from each horizontal curve produce better results as the vertical separation D increases. Depth differences from the middle curve produce a higher R_1RR than those from the top and bottom curves, yielding the highest R_1RR of 86.03% when D_2 = 20 for the FRGC database and 92.37% when D_2 = 22 for the Bosphorus database. The R_1RR of depth differences from the top curve increases up to $D_1 \approx 15$ and becomes more stable afterwards. In contrast, the R_1RR of depth differences from the bottom curve increase more slowly and when $D_3 \geq 10$ it becomes relatively constant, with some fluctuations.

An improved performance can be achieved by combining the depth differences from all three curves. However, as the vertical separation D increases the depth differences from the top and bottom curves are more affected by natural occlusions (e.g. glasses and hands) and shape changes caused by expression, respectively, and so simply combining the best performing distances from each curve will not necessarily produce the best combined performance. An evaluation of the combined depth difference features on the FRGC and Bosphorus databases using different distances for D_1, D_2 and D_3 was undertaken and the best performance was found to occur when D_1 and D_3 are ≈ 15 and D_2 is ≈ 20.

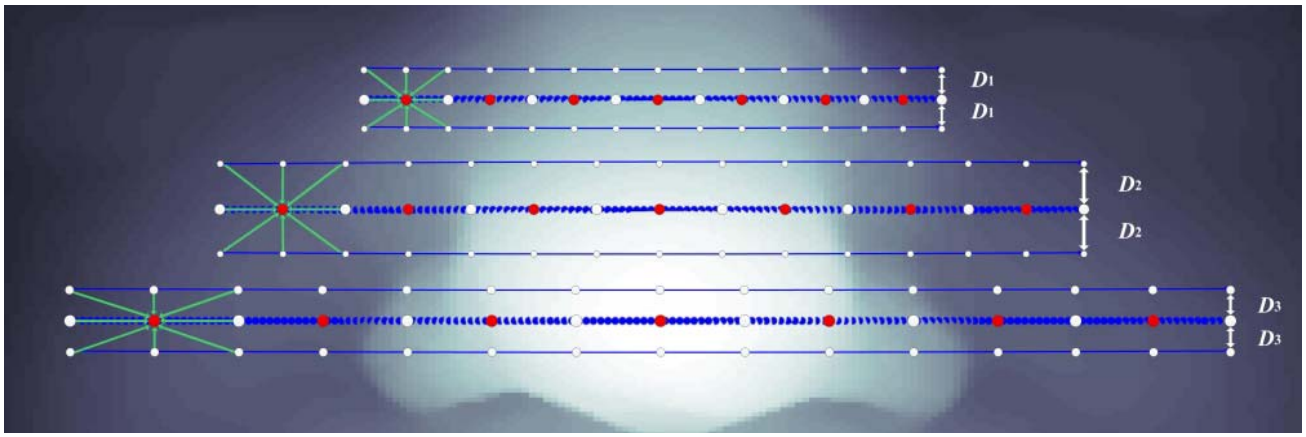
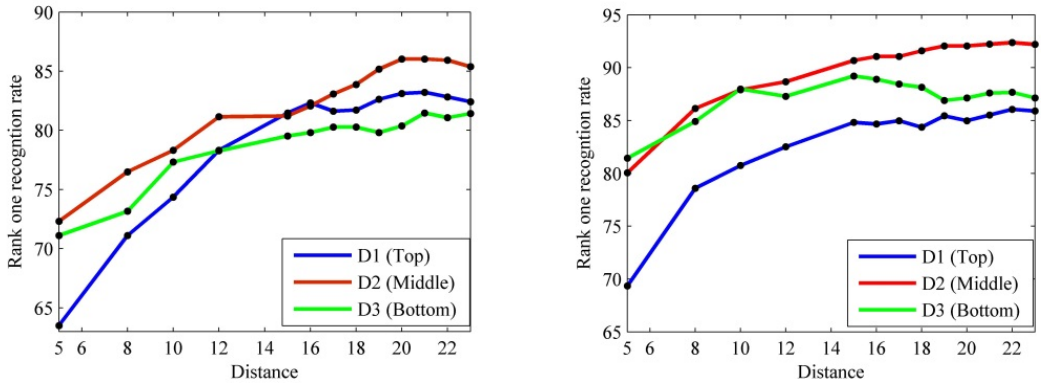
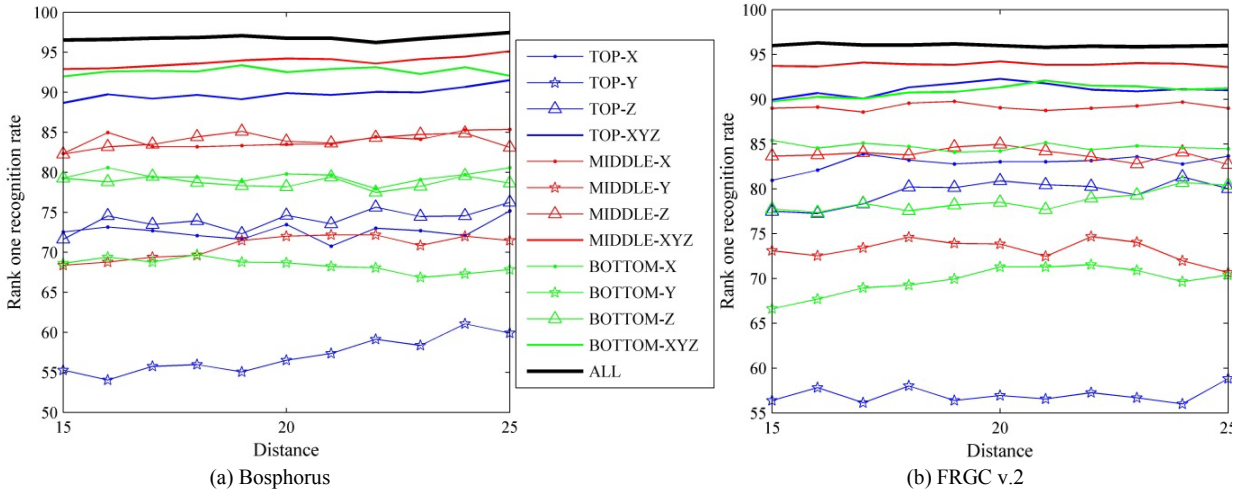


Fig. 3. Local features extraction by calculating the depth difference around the three horizontal curves.



(a) FRGC (b) Bosphorus

Fig. 4. Rank one recognition rate of depth difference extracting around the three horizontal curves under different distances on the FRGC database.



(a) Bosphorus (b) FRGC v.2

Fig. 5. Rank one recognition rate of features extracting from horizonal curves with different separation distances using three channels (x, y and z) and three parts of nose and adjoining regions on the Bosphorus and FRGC v.2 database.

Using these parameters, the recognition performance is evaluated under both the identification and verification scenarios and the results are presented in TABLE II. The performance for captures with expression variations is significantly better than those achieved using the large scale depth difference (see TABLE I), albeit with a larger feature set.

In addition to depth difference calculation, the local difference descriptor is also applied to the normal component images to find the discriminative features from the *x*, *y* and *z* components, producing another three different representations of this region. The scale of the windows for the normals was the same as that used for depth and the recognition performance for both databases is illustrated in TABLE II.

Normal features outperform the depth ones on the Bosphorus database, which contains many variations in expression, and also for the FRGC "All vs Non-neutral" scenario. Results also show that the R_1RR for the Bosphorus database is much higher than for FRGC experiment 3. This is most likely due to the noisy training captures in the Spring 2003 folder.

In addition to the analysis of the recognition performance for the depth component in Fig. 4, some additional identification evaluations are used to further explore the discriminatory power of different locations on the selected region (Top, Middle and Bottom) and three surface normal components (*x*, *y* and *z*) using a fixed scale, see TABLE III. The R_1RR results are for the Bosphorus database, experiment 3 of FRGC and the FRGC v.2 databases. In general, the results in TABLE III show that *x* and *z* possess more discriminative features than *y* and produce a performance that is ~4% better. For the horizontal regions of cheek/nose region, both the depth and normal features extracted from Middle part outperform those from the Top and Bottom parts.

An additional evaluation on the relationship between the R_1RR and different scales is shown in Fig. 5 where horizontal curves with separations ranging from 15 to 25 pixels are investigated. The results show that the R_1RR remains stable when different separations are selected, proving that the proposed local shape descriptor is scale robust to some extent. Also, *x*, *z* and Middle components always produce better R_1RR, regardless of which scale is selected.

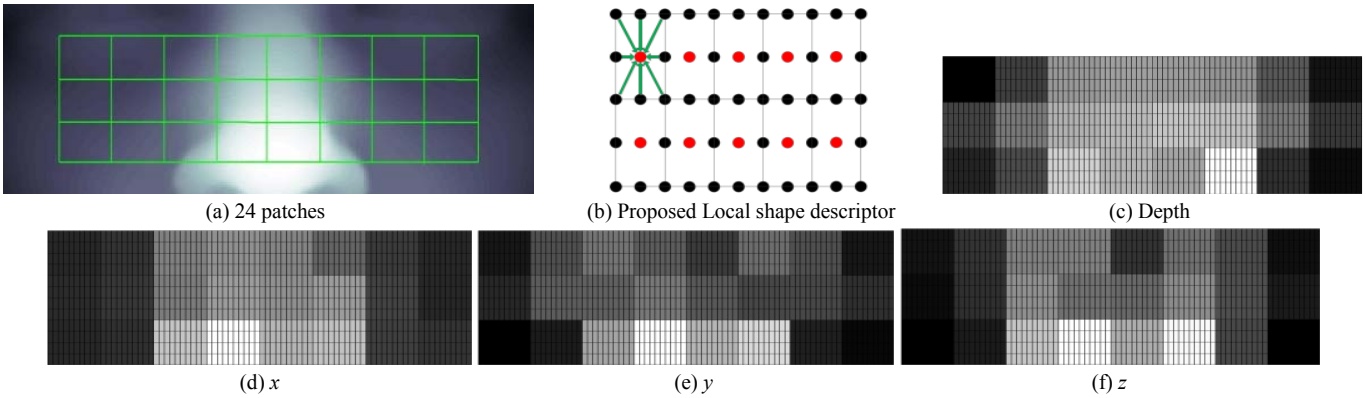
The results for the whole FRGC database shown in the TABLEs II and III reveal that surface normals are much more sensitive to the noise caused by lighting and the capture device. Training data from noisy folder, Spring 2003, affects the intra-class similarity and inter-class dissimilarity and so cannot reliably be used to select the most discriminative normal features. To investigate this, Fig. 5 evaluates the normal features on the FRGC v.2 database, 466 subjects with 4007 captures in total, and presents the R_1RR of features selected using different distances with three channels (*x*, *y* and *z*) and three parts (top, middle and bottom). The recognition performance achieved is quite similar to that of the Bosphorus database, showing the adverse effect of the noisy Spring 2003 captures on the recognition performance.

TABLE II. RECOGNITION PERFORMANCE USING DEPTH AND SURFACE NORMAL DIFFERENCES AROUND THE THREE CURVES IN FIG. 3.

	Scenarios	Depth Features		Normal Features	
		R_1RR	EER	R_1RR	EER
Bosphorus [13]	All vs All	94.99%	3.47%	97.76%	1.32%
Experiment 3 in FRGC [12]	All vs All	89.63%	4.51%	89.08%	3.90%
	All vs Neutral	93.38%	3.03%	91.67%	3.11%
	All vs Non-neutral	82.89%	5.88%	84.13%	4.91%
Features size of each capture		8×7×3 = 168		8×7×3×3= 504	

TABLE III. RECOGNITION PERFORMANCE OF FEATURES EXTRACTING FROM THREE CHANNELS (*X*, *Y* AND *Z*) AND THREE CURVES USING A FIXED SCALE

	Components	Size	Bosphorus[13]	Experiment 3 in FRGC [12]	FRGC v.2 [12]
Surface Normal	X	168	94.60%	83.62%	94.72%
	Y	168	90.28%	78.42%	90.19%
	Z	168	94.83%	83.78%	92.77%
	Top	168	91.52%	81.87%	92.26%
	Middle	168	95.14%	83.93%	94.21%
	Bottom	168	93.37%	79.36%	92.08%
	All	504	97.76%	89.08%	96.29%
Depth	Top	56	84.81%	83.22%	84.16%
	Middle	56	92.06%	86.03%	91.39%
	Bottom	56	87.12%	81.47%	87.12%
	All	168	94.99%	89.63%	96.17%



(a) 24 patches (b) Proposed Local shape descriptor (c) Depth

(d) *x* (e) *y* (f) *z*

Fig. 6. The patch weights corresponding to depth image and three normal component images learned from the Bosphorus database. The weights are presented as gray scale values, where brighter values denote larger weights.

IV. Local Patches Evaluation on the Nasal and Its Environs

The proposed local shape descriptor demonstrates good potential for extracting expression robust features on a larger scale with a small sized feature set, using the nose and adjoining cheek regions as an expression robust biometric. It is therefore interesting to use this descriptor to explore the discriminatory power more locally.

A. Discriminatory Power Evaluation on Local Patches

The nasal and adjoining cheek regions shown in Fig. 6(a) are used to further investigate the discriminative features for expression robust face recognition, which is determined by two cheek landmarks, and the top and bottom horizontal curves shown in Fig. 1. This is an extension of the region discussed above, which is further divided into 24 small local patches, each of which is resampled to 5×11 pixels, as illustrated in Fig. 6(b). Calculating the local shape descriptor proposed in Section III(A) builds the feature set from both the depth and normal components. The R_1RR of each patch is obtained using the KFA classifier and cosine distance. To illustrate the discriminatory power, the R_1RRs for the patches are presented as grayscale images in Fig. 6(c) - (f). Results show that the R_1RR of patches on the nasal region and its adjoining cheek regions is always higher than those on the cheek bone region, for all four components.

Furthermore, it is more interesting to observe from Fig. 6 that those patches located on the boundary of the nasal and adjoining cheek regions perform better than those only wholly in the nasal region which shows the potential of the cheek regions immediately adjoining the nose. The features were also evaluated on the FRGC v.2 database, producing similar recognition performances and discriminatory power distribution for the depth and three normal components as that shown in Fig. 6. This analysis shows that the grayscale images in Fig. 6 can be considered to be a good representation of the discriminatory power on the nasal and adjoining cheek regions.

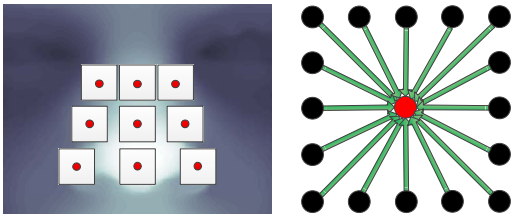
B. Square Local Shape Descriptor

The nine landmarks (red points) shown in Fig. 7(a) are localized accurately to define the nasal and its environs region. The two nasal alar landmarks are found by intersecting a set of planes which are perpendicular to the *xy* plane. This results in a set of curves and the alar landmarks are located at the minimum of the curves' maxima [2]. The vertical middle three landmarks are determined by the location of saddle and tip and are evenly distributed on the line from tip to saddle. Similarly, the three squares on the right and left three are located using the two alar and saddle landmarks.

The nine square windows centered on the nine landmarks are shown in Fig. 7(a). The edge of each window can be resampled to a series of points and the local features are found by calculating the differences between those edge points and central point in each window, which is an extension of the proposed local shape descriptor presented in Section III using the following equation:

$$\begin{cases} F_{Depth} = \sum_{j}^{n}\sum_{i}^{m}(E_{Depth}(i) - C_{Depth}(j)) \\ F_X = \sum_{j}^{n}\sum_{i}^{m}(E_X(i) - C_X(j)) \\ F_Y = \sum_{j}^{n}\sum_{i}^{m}(E_Y(i) - C_Y(j)) \\ F_Z = \sum_{j}^{n}\sum_{i}^{m}(E_Z(i) - C_Z(j)) \end{cases} \quad (1)$$

where n is the number of windows, m is the number of points on the edge, E refers to the value of each edge points and C denotes the central points. In this experiment, each of the nine windows has 16 edge points resulting in feature sets of size 9×16 = 144 for features extracted from each components. F_{Depth} is depth component and F_X, F_Y and F_Z are the x, y and z normal components.



(a) (b)

Fig. 7. 9 landmarks and their corresponding patches on the nasal region

C. Recognition Performance Evaluation Using Different Scales

The influence of the scale of the window on the recognition performance can be evaluated by altering the vertical distance from center to edge of the square patch. To this end, sizes ranging from 4 to 26 are used to evaluate the recognition performance using the Bosphorus database under two identification scenarios: the KFA classifier using cosine distance and the leave-one-out (LOO) classifier using the city block distance.

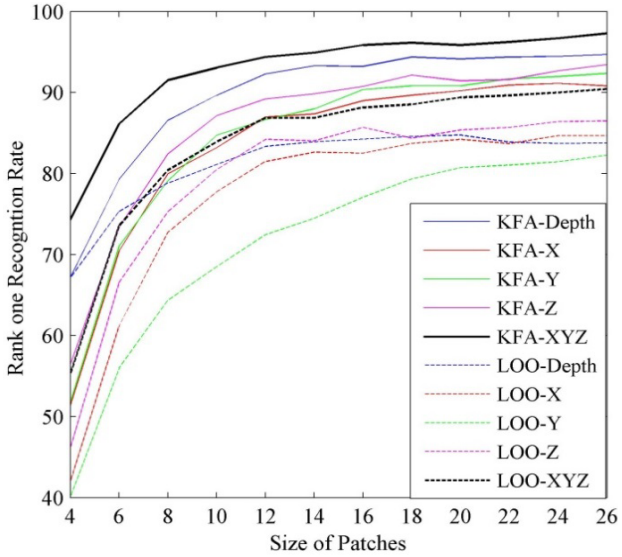


Fig. 8. Rank one recognition rate calculated by both KFA and LOO classifier using different size of windows (patches) on the Bosphorus database.

The LOO scenario employs every capture as a probe against a gallery of all other captures from different identities. The benefit of using LOO is that all sessions are used for the probe and gallery captures, which making it a far more efficient scenario than using a fixed partition. The L1-norm [16], city block distance is used to compare the probes and gallery captures as it has better discriminatory power when the feature space is sparse.

The R_1RRs for the KFA and LOO classifiers using different patch sizes are shown in Fig. 8. Although the R_1RR improves with increasing scale for all components, it remains stable with no significant improvement after the patch size reaches 12. Thus 12 is a good compromise scale, as it covers the whole nasal and adjoining regions without the risk of potential occlusions that affects larger patch sizes.

In addition, on the basis of the general analysis of discriminatory power map of 24 patches, the recognition performance of 9 patches shows a high potential for extracting expression robust discriminative features from the nasal and adjoining regions. The experiments results demonstrate that these smaller sized features can produce a comparable recognition performance.

V. Discussion and Conclusions

In this paper, the nasal and adjoining cheek regions have been explored as an expression robust biometric for 3D face recognition. Focusing on these regions is advantageous as they have been shown to be more invariant to variations due to universal expressions and natural occlusions. Thus, extracting features from these stable patches on the nasal and its environs has a high potential to build a more powerful 3D face recognition algorithm.

In addition, a novel 3D face local shape representation method is proposed to provide a simple way to generate a small feature set that produces high recognition performance. A set of features from the 3D face are found by calculating the depth differences between central points and eight neighboring points from horizontal curves on the face surface. Experiments show that using the depth differences from three horizontal curves as a feature vector produces a better recognition performance than using the depth values on the curves directly, providing motivation for further work. More localized depth and normal differences were found by including additional points from above and below each horizontal curve and experiments for both the identification and verification scenarios demonstrated that a good performance can be achieved with a relatively small feature set.

Using the proposed local shape descriptor a further analysis on the discriminatory power of 24 patches on the nose and adjoining cheek regions was evaluated, producing a set of novel discriminative maps for both depth and normals (x, y and z). On the basis of the resulting discriminative maps, an extension of the new local shape descriptor was applied to the nine patches on the nasal and adjoining cheek regions. Results showed that the work has potential for use as an expression robust 3D face biometric or possibly a pattern rejector.

A series of experiments have been presented to evaluate the recognition performance of the proposed multi-component and scale local shape descriptor on the nasal and adjoining cheek regions, focusing on addressing the problem of expression variations on the human face. Using a simple descriptor to extract features results in a very small feature set, which has much potential for real time automatic face recognition system as all the preprocessing parts (face cropping, alignment and landmark localization) are fully automated and no feature selection is required after the feature extraction.

For classification and matching, the KFA classifier with the cosine distance produces a R_1RR of 97.76% and an EER of 1.32% for the Bosphorus dataset, which is competitive with methods whose algorithms are more sophisticated and require a larger feature set. Comparison with other research that considers the nasal region and its environs, see TABLEs IV and V, shows that the proposed method has much potential. For verification, using the FRGC database (Experiment 3), the EER for both depth and normals was found to be ~3% for neutral and ~5% for non-neutral which compares favorably with the results of Chang *et al*. [1] (12% and 23%, respectively) and Emambakhsh *et al*. [2] (8% and 18%, respectively). For identification, compared to the algorithms using both selected region and the whole face in TABLEs V and VI, the combined surface normals results exceed 96.29% and 97.76% on the FRGC v.2 and Bosphorus databases, respectively, with a relatively small and simple feature set which is within 2% of the state-of-the-art performance using the whole face.

TABLE IV. Comparison of EERs tested on the FRGC database (Experiment 3) using nasal region and its environs

Algorithm	Matching	Neutral	Non-neutral
Chang *et al*. [1]	ICP	0.12	0.23
Emambakhsh *et al*. [2]	KFA-Poly	0.08	0.18
Gao *et al*. [7]	KFA-Poly	0.03	0.06
Proposed method	KFA-Poly	0.03	0.05

TABLE V. Comparison of R_1RRs tested in the FRGC (Experiment 3) and Bosphorus Databases using nasal region and its environs

Algorithm	FRGC		Bosphorus
	Neutral	Non-neutral	
Chang *et al*. [1]	96.6%	82.7%	
Emambakhsh *et al*. [2]	90.87%	81.61%	97.44%
Gao *et al*. [7]	92.68%	75.32%	93.61%
Proposed method	93.38%	84.13%	97.76%

TABLE VI. Comparison of R_1RRs tested in the FRGC (Experiment 3) and Bosphorus Databases using nasal region and its environs

	Algorithm	FRGC		Bosphorus
		Experiment 3	v.2	
Nasal and its environs	**Gao *et al*. [7]**	86.33%		93.61%
	Wang *et al*. [4]		~92%	
	This work	89.63%	96.29%	97.76%
The whole face	**Wang *et al*. [4]**		98.3%	
	Li *et al*. [10]		96.3%	95.4%
	Kakadiaris *et al*. [17]		97.9%	98.2%
	Ballihi *et al*. [6]		98.02%	
	Drira *et al*. [8]		97%	
	Faltemier *et al*. [18]		97.2%	

This paper has not evaluated the single training sample scenario. Therefore, one area of future work will focus on exploiting an effective subset from the original features using some feature selection methods or fusing other features to build a new feature vector with more discriminative information for matching.

ACKNOWLEDGMENT

Jiangning Gao gratefully acknowledges her financial support from the China Scholarship Council (CSC) and a University of Bath Graduate School Scholarship.

REFERENCES

[1] K. I. Chang, W. Bowyer, and P. J. Flynn, "Multiple Nose Region Matching for 3D Face Recognition under Varying Facial Expression," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 28, pp. 1695-1700, 2006.

[2] M. Emambakhsh, A. Evans, and M. Smith, "Using nasal curves matching for expression robust 3D nose recognition," in Proc. Biometrics: Theory, Applications and Systems, 2013.

[3] A. S. Mian, M. Bennamoun, and R. Owens, "An Efficient Multimodal 2D-3D Hybrid Approach to Automatic Face Recognition," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 29, pp. 1927-1943, 2007.

[4] W. Yueming, T. Xiaoou, L. Jianzhuang, P. Gang, and X. Rong, "3D Face Recognition by Local Shape Difference Boosting," in Proc. European Conference on Computer Vision, 603-616, 2008.

[5] M. Emambakhsh and A. Evans, "Self-dependent 3D face rotational alignment using the nose region," in Proc. 4th International Conference on Imaging for Crime Detection and Prevention, London, 2011.

[6] L. Ballihi, B. Ben Amor, M. Daoudi, A. Srivastava, and D. Aboutajdine, "Boosting 3-D-Geometric Features for Efficient Face Recognition and Gender Classification," IEEE Transactions on Information Forensics and Security, vol. 7, pp. 1766-1779, 2012.

[7] J. Gao, M. Emambakhsh, and A. N. Evans, "A Low Dimensionality Expression Robust Rejector for 3D Face Recognition," in Proc. IEEE International Conference on Pattern Recognition, Sweden, pp. 506-511, 2014.

[8] H. Drira, B. Ben Amor, A. Srivastava, M. Daoudi, and R. Slama, "3D Face Recognition Under Expressions, Occlusions and Pose Variations," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 35, pp. 2270 - 2283, 2013.

[9] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24, pp. 971-987, 2002.

[10] H. Li, D. Huang, J.-M. Morvan, L. Chen, and Y. Wang, "Expression-robust 3D face recognition via weighted sparse representation of multi-scale and multi-component local normal patterns," Neurocomputing, vol. 133, pp. 179-193, 6/10/ 2014.

[11] S. Zafeiriou, G. A. Atkinson, M. F. Hansen, W. A. P. Smith, V. Argyriou, M. Petrou, et al., "Face Recognition and Verification using Photometric Stereo: The Photoface Database and a Comprehensive Evaluation," IEEE Transactions on Information Forensics and Security , vol. 8(1), pp. 121-135, 2012.

[12] P. J. Phillips, P. J. Flynn, T. Scruggs, K. W. Bowyer, C. Jin, K. Hoffman, et al., "Overview of the face recognition grand challenge," in Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 1, pp. 947-954, 2005.

[13] A. Savran, N. Alyüz, H. Dibeklioğlu, O. Çeliktutan, B. Gökberk, B. Sankur and L. Akarun, "Bosphorus Database for 3D Face Analysis," in Biometrics and Identity Management, Springer Berlin Heidelberg, pp. 47-56, 2008.

[14] D. Smeets, P. Claes, J. Hermans, D. Vandermeulen, and P. Suetens, "A Comparative Study of 3-D Face Recognition Under Expression Variations," IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews, vol. 42, pp. 710-727, 2012.

[15] C. Liu, "Capitalize on dimensionality increasing techniques for improving face recognition grand challenge performance," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 28, pp. 725-737, 2006.

[16] J. Wright, A. Yang, A. Ganesh, S. Sastry, and Y. Ma. "Robust face recognition via sparse representation," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 31(2), pp. 210–227, 2009.

[17] O. Ocegueda, G. Passalis, T. Theoharis, S. Shah, I. Kakadiaris, "Ur3d-c: linear dimensionality reduction for efficient 3d face recognition," in Proc. IEEE International Joint Conference on Biometrics, 2011.

[18] T.C.T. Faltemier, K.W.K. Bowyer, and P.J.P. Flynn, "A Region Ensemble for 3D Face Recognition," IEEE Transactions on Information Forensics and Security, vol. 3, no. 1, pp. 62-73, Mar. 2008.